EPA
United States
Environmental Protection
Agency

Forecasting Exposure in Order to Use High Throughput Hazard Data in a Risk-based Context

John Wambaugh
National Center for Computational Toxicology

9th World Congress on Alternatives and Animal Use in the Life Sciences
24-28 August 2014
Prague, Czech Republic

Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Introduction

- The timely characterization of the human and ecological risk posed by thousands of existing and emerging commercial chemicals is a critical challenge facing EPA in its mission to protect public health and the environment

- ExpoCast is an initiative to develop the necessary approaches and tools for rapidly predicting exposure for thousands of chemicals (Cohen-Hubal, *et al.*, 2010)
- **Proof of Concept:** Used off-the-shelf high throughput exposure models and evaluated predictions with biomonitoring data to characterize uncertainty (Wambaugh *et al.,* 2013)

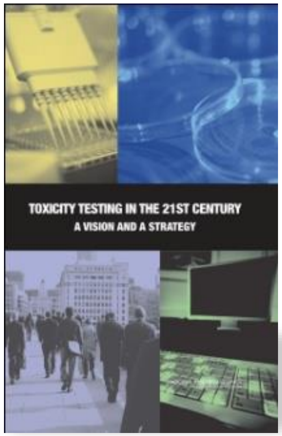
Environmental Fate and Transport

Consumer Use and Indoor Exposure

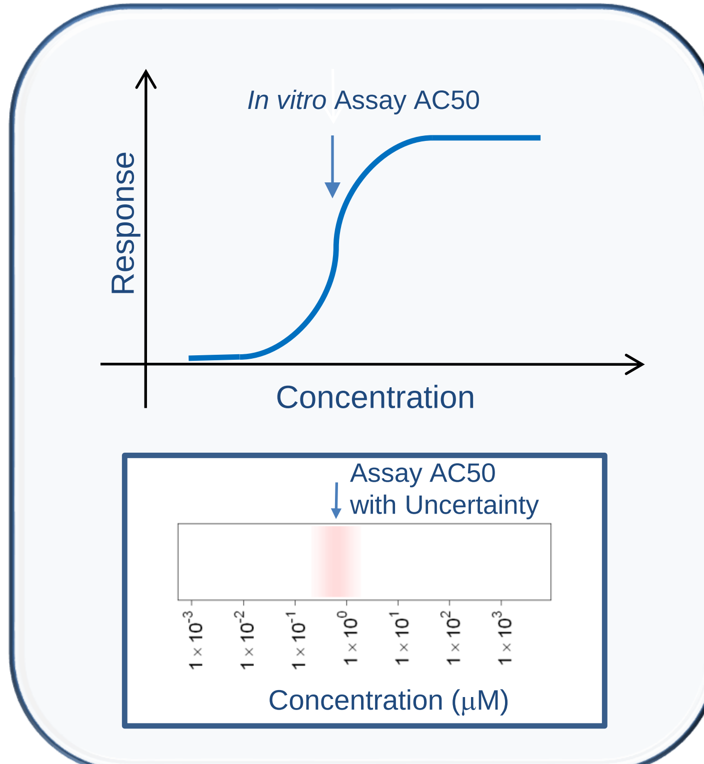
“All cases are unique, and very similar to

United States Environmental Protection Agency

High-Throughput Toxicity Testing



Tox21: Examining ~10,000 chemicals using ~50 assays intended to identify interactions with biological pathways (Schmidt, 2009)

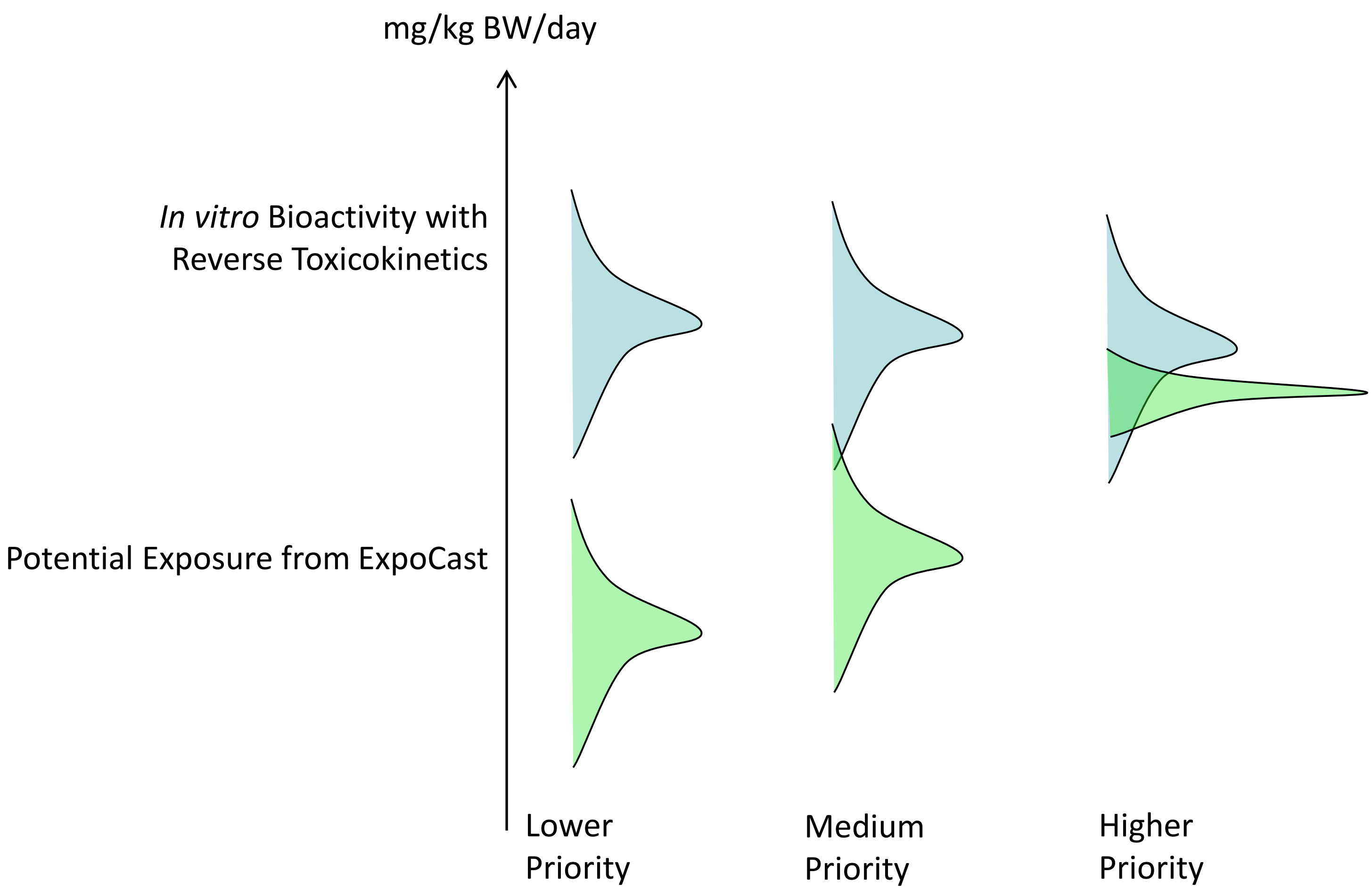


ToxCast: For a subset (>1000) of Tox21 chemicals ran >500 additional assays (Judson *et al.,* 2010)

Most assays conducted in dose-response format (identify 50% activity concentration – AC50 – and efficacy if data described by a Hill function)

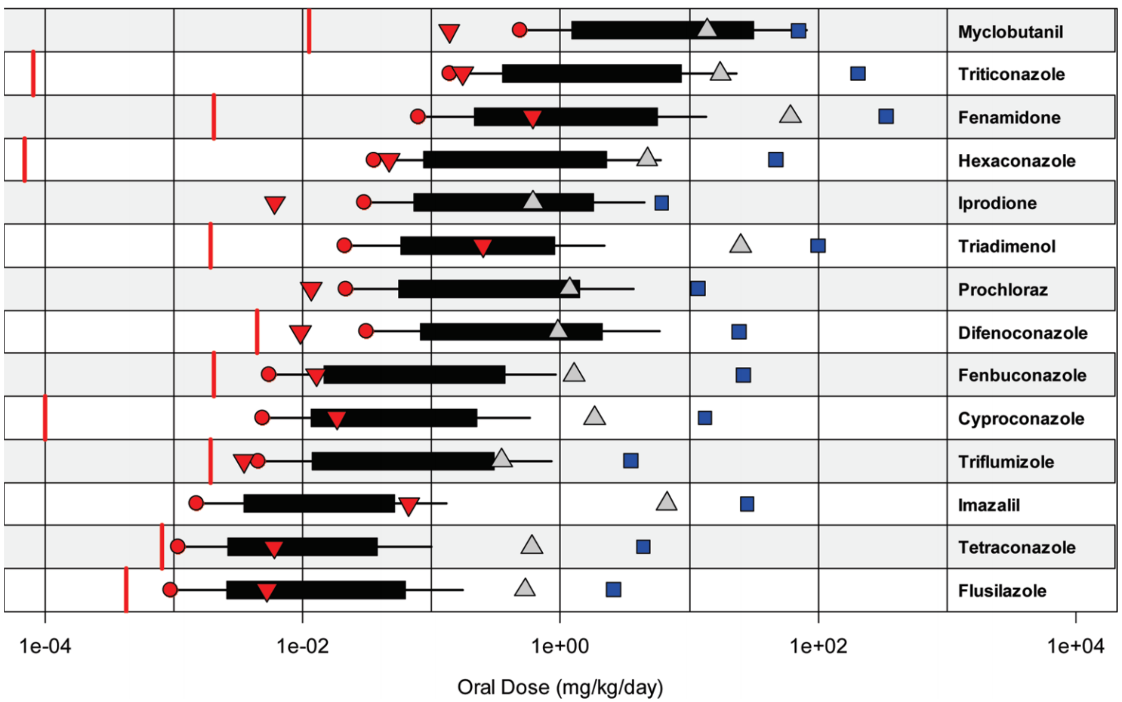
All data is public: http://actor.epa.gov/

Application of RTK to EDSP



e.g. Judson *et al.,* (2011)

In vitro Bioactivity, RTK and *in Vivo* Toxic Doses



Comparison oral equivalent doses (box and whisker plots in mg/kg/day) predicted with RTK and LEL and NEL values for liver hypertrophy from animal studies

- ■ **Lowest Observed Effect Level**
- △ **No Observed Effect Level (NEL)**
- ▼ **NEL/100**

Estimated chronic exposure levels from food residues are indicated by vertical red lines. All values are in mg/kg/day.

Figure from Judson *et al*. (2011)

EPA
United States
Environmental Protection
Agency

Exposure Science in the 21st Century

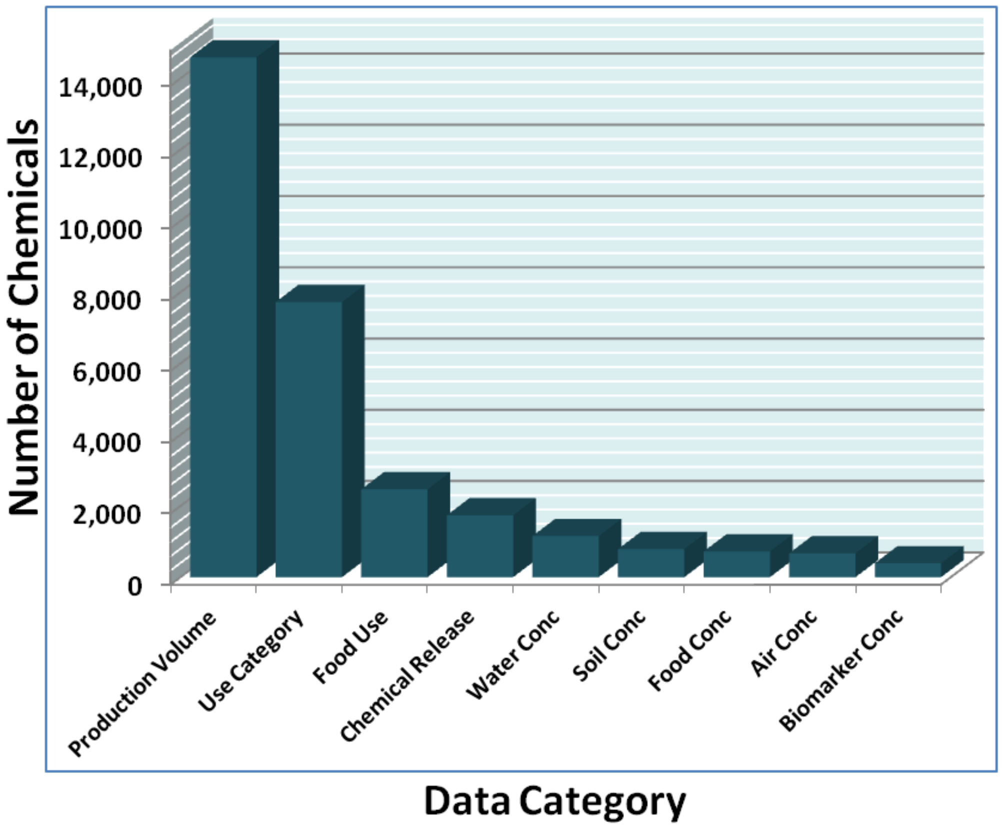
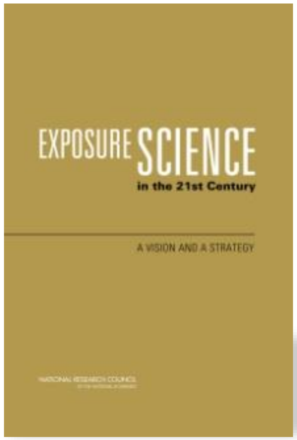


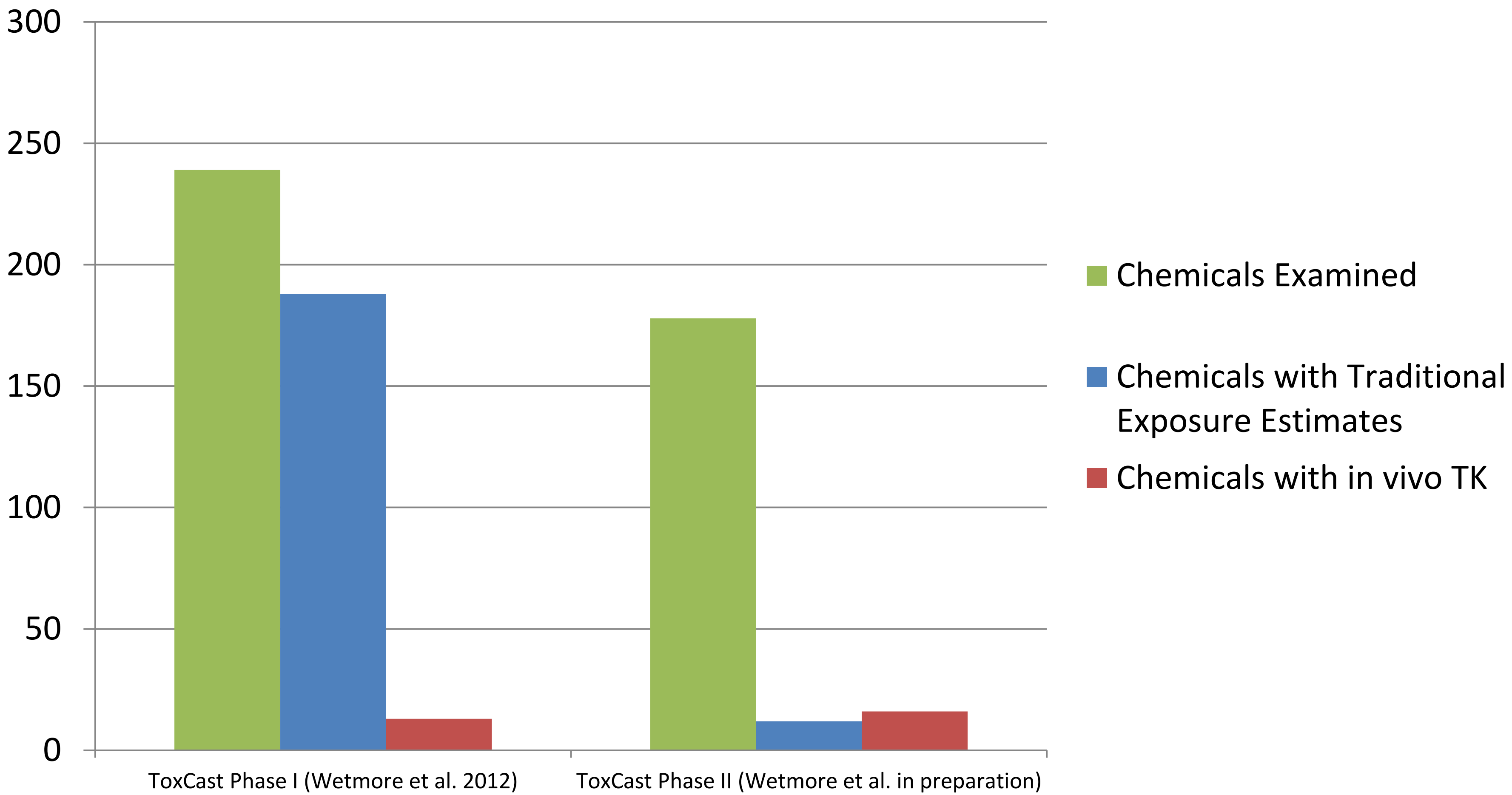
Figure from Egeghy et al. (2012),
"The exposure data landscape for manufactured chemicals"

- 2012 NRC report:



  - New tools needed for screening and prioritization of chemicals for targeted toxicity testing
  - New, focused exposure assessments or monitoring studies needed
  - Better quantification of population vulnerability needed

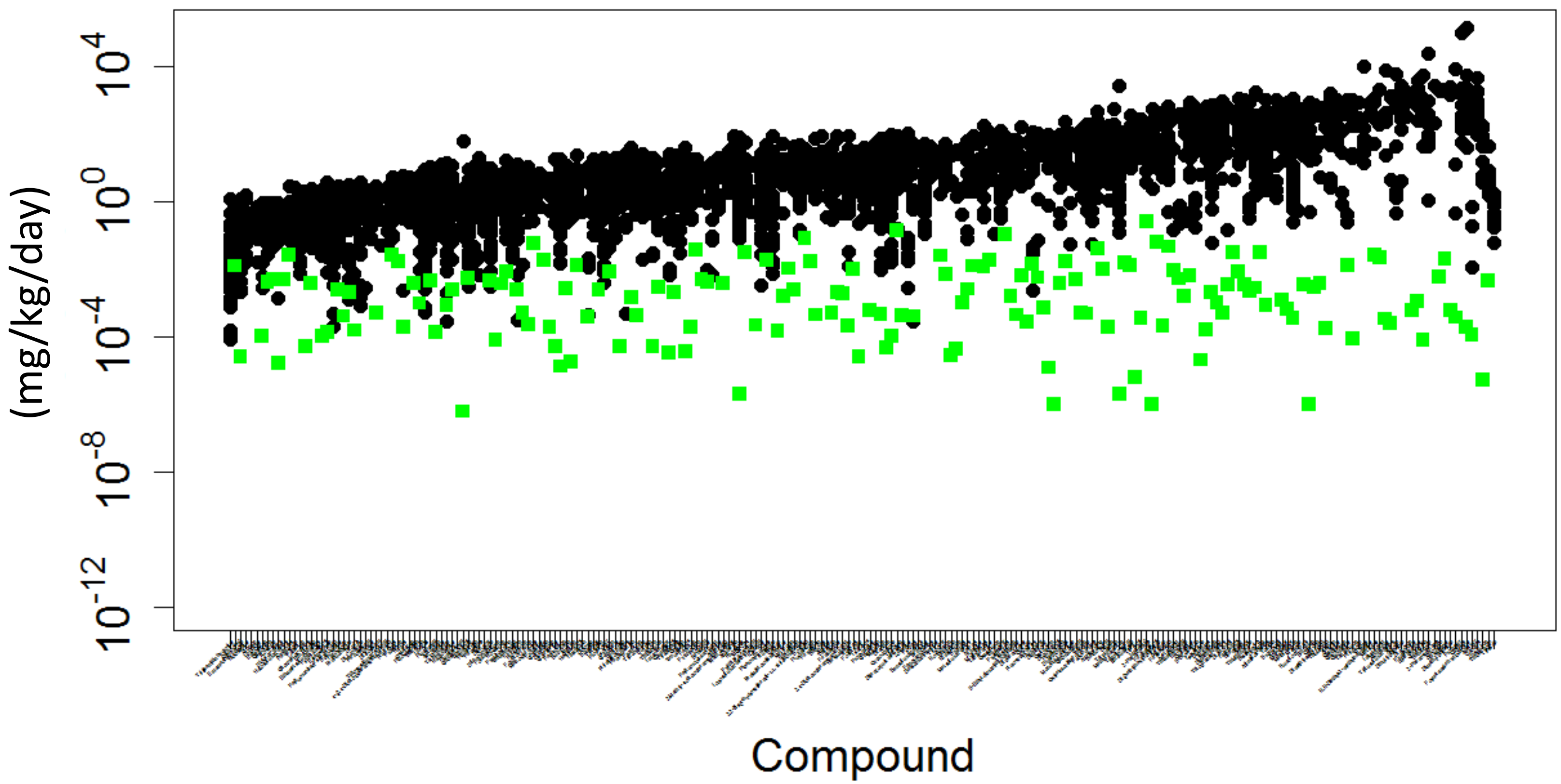
In Vitro Bioactivity, *In Vivo* Pharmacokinetics, and Exposure



- As in Egeghy et al. (2012), there is a paucity of data for providing context to HTS data

ToxCast Phase I Oral Equivalent Doses and Exposure Estimates

ToxCast Phase I chemicals included many pesticide active ingredients

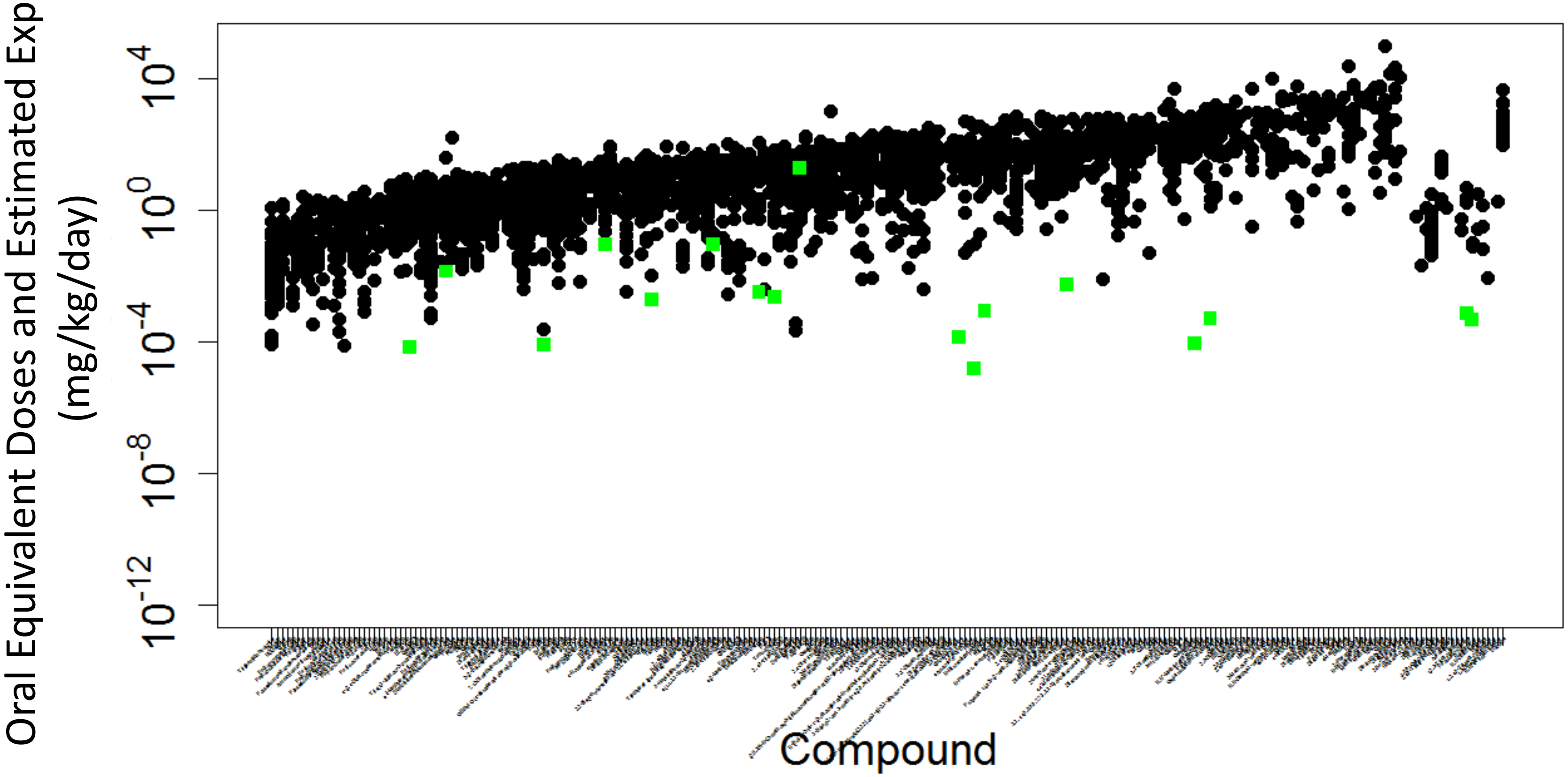


Green squares indicate highest estimated exposures from EPA REDs or CDC NHANES: ~71% of Phase I

Wetmore et al. *Tox. Sci (2012)*

The Exposure Coverage of the ToxCast Phase II Chemicals

ToxCast Phase II: mostly non-pesticides, including plasticizers and pharmaceuticals



Green squares indicate highest estimated exposures from EPA REDs or CDC NHANES: ~71% of Phase I

~7% of Phase II

Unpublished data from Barbara Wetmore

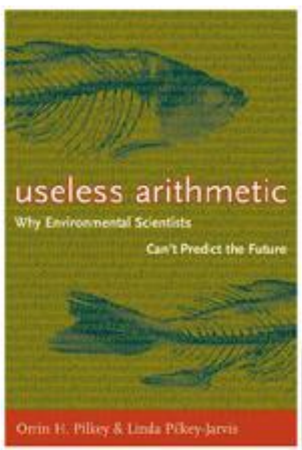
United States Environmental Protection Agency

How to Make Good Forecasts

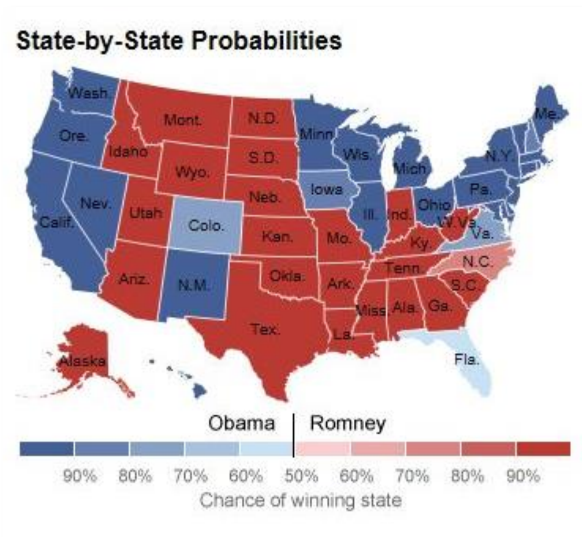
1) Think probabilistically: ExpoCast evaluates model performance systematically across as many chemicals (and chemistries) as possible

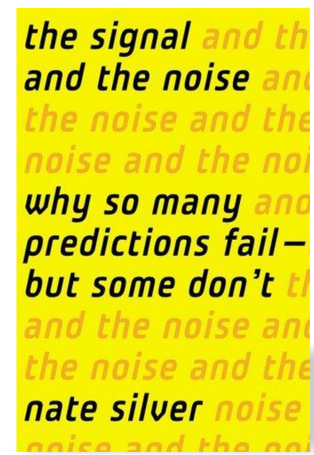
2) Forecasts change : Today's forecast reflects the best available data today but we must accept that new data and new models will cause predictions to be revised

3) Look for consensus: We evaluate as many models and predictors/predictions as possible



Orrin Pilkey &
Olinda Pilkey-Jarvis (2007)





Nate Silver (2012)

Exposure Forecasting (ExpoCast)

- Develop the tools and data necessary to rapidly quantify human and ecological exposure potential of chemicals

- Focus is distinct from many existing exposure tools that support either screening level assessments on a per chemical basis or full regulatory risk assessment

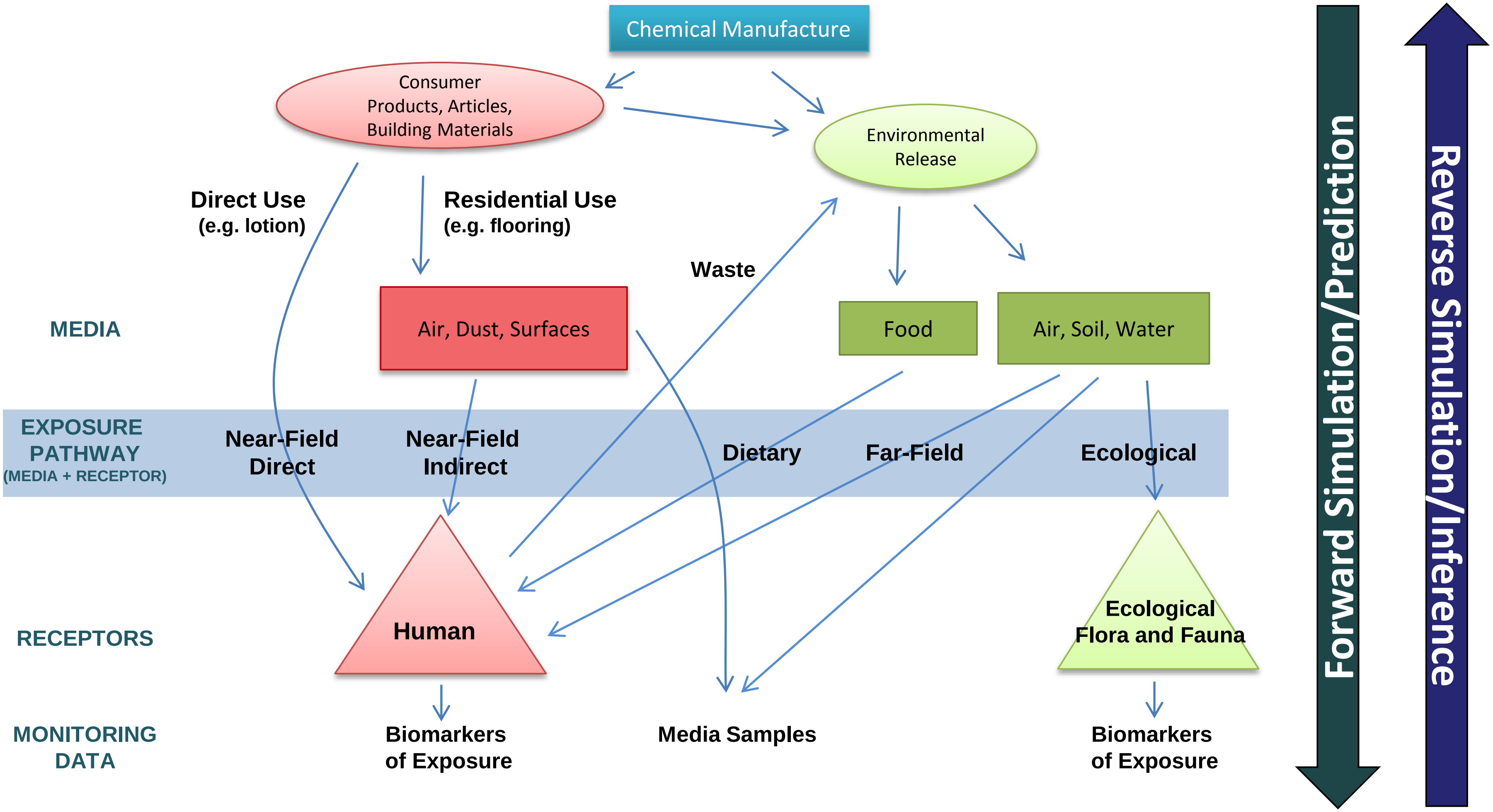
In Nate Silver's terminology:

a ***prediction*** is a specific statement

a ***forecast*** is a probabilistic statement

Wikipedia (statistics): "when information is transferred across time, often to specific points in time, the process is known as forecasting"

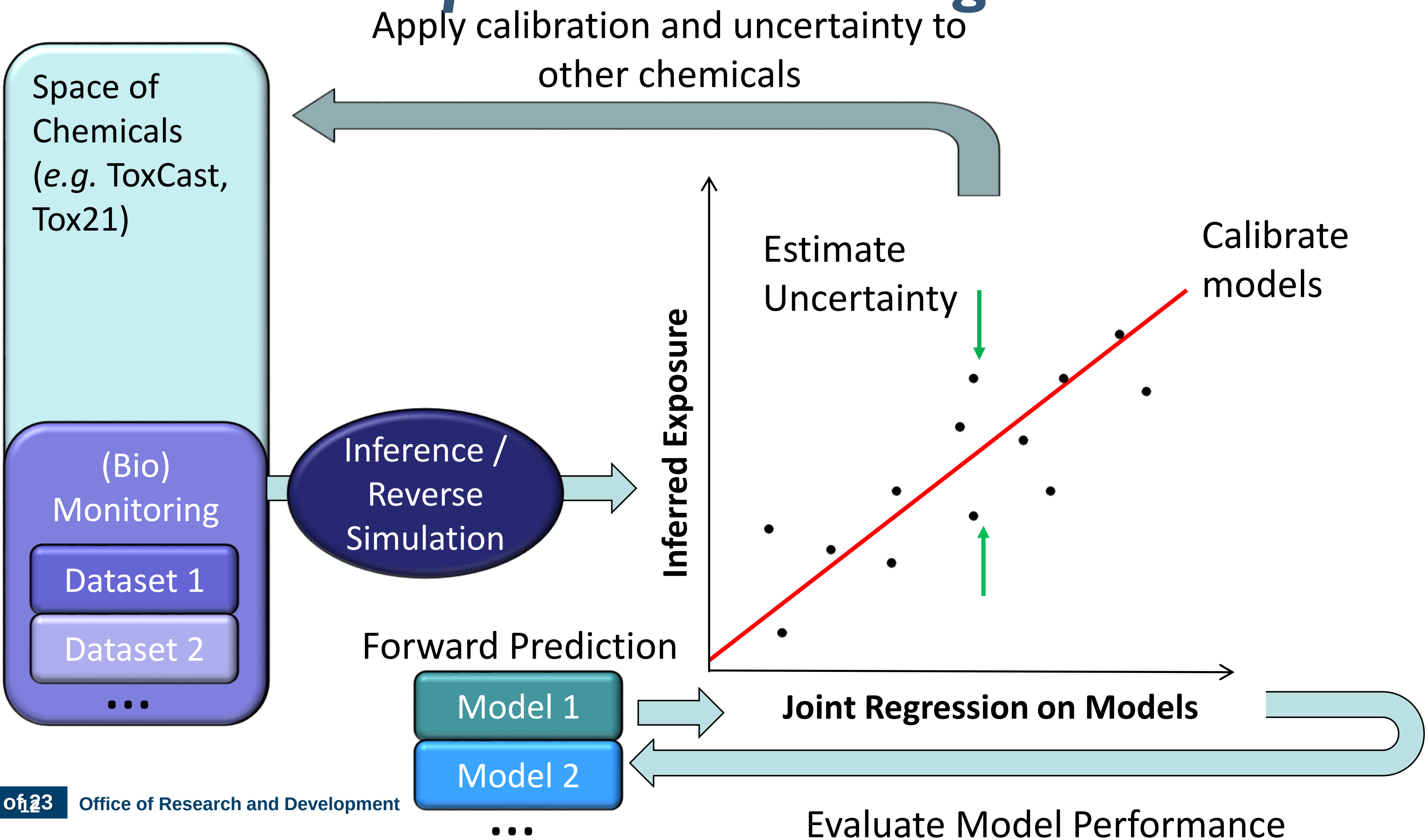
Investigating Exposure



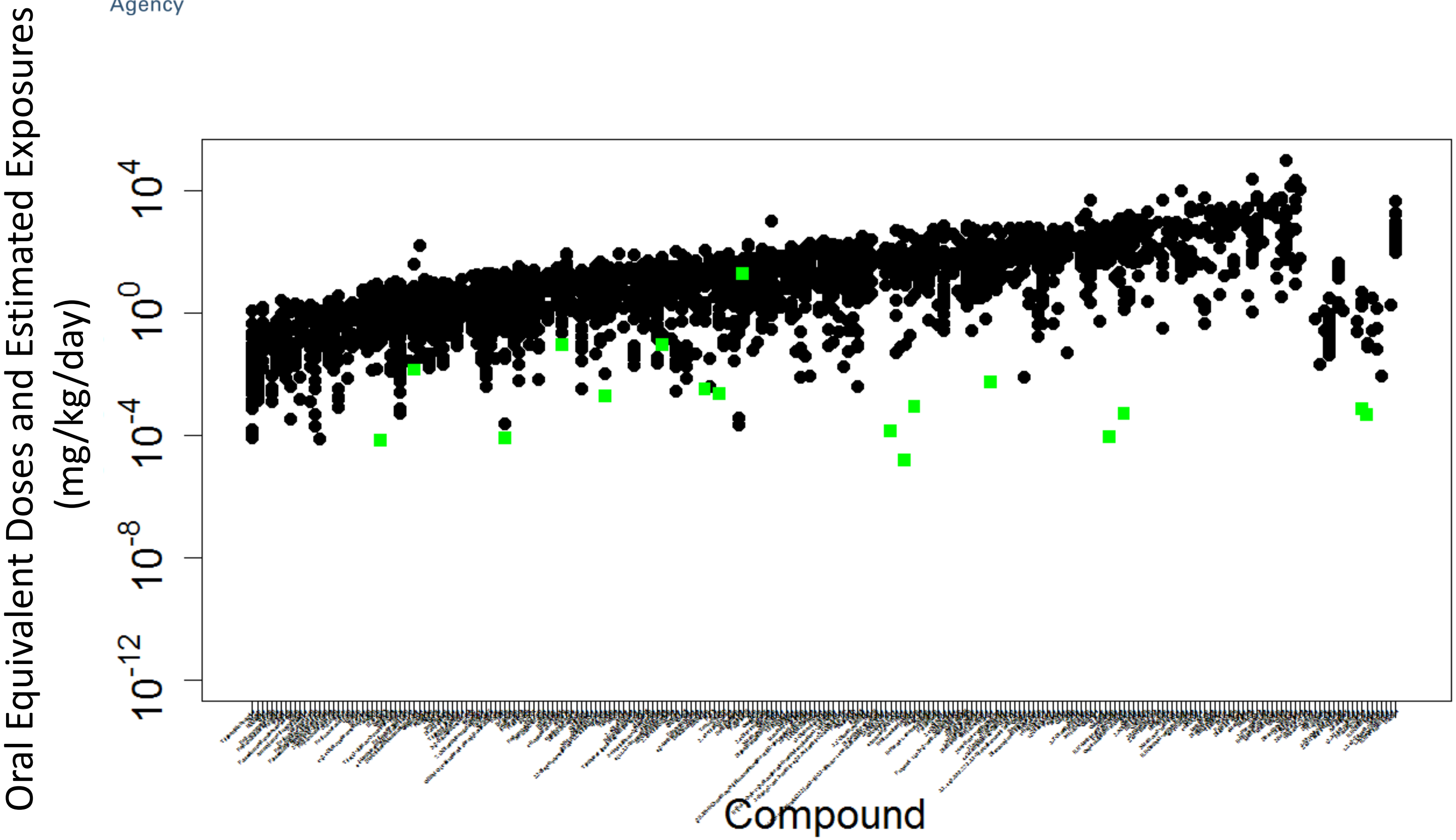
 Office of Research and Development

Figure modified from Kristin Isaacs original

Framework for High Throughput Exposure Screening



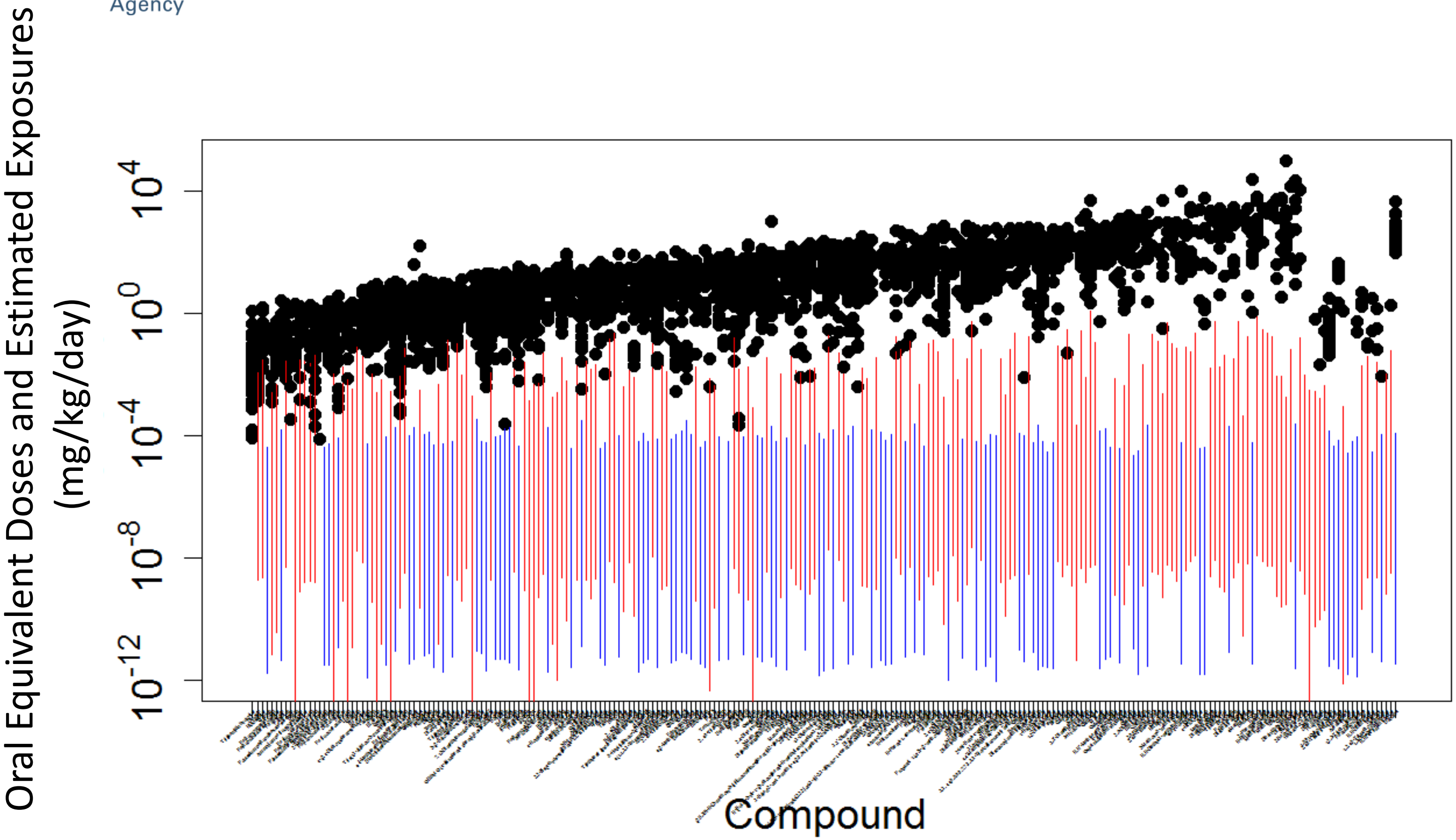
The Exposure Coverage of the ToxCast Phase II Chemicals



Green squares indicate chemicals with any sort of exposure estimate available.

Unpublished data from Barbara Wetmore

ExpoCast Coverage of the ToxCast Phase II Chemicals



Predictions from Wambaugh *et al.* (2013) ExpoCast model with USEtox, RAIDAR, and **near field**/**far field** heuristic

ToxCast Oral Equivalents based on unpublished data from Barbara Wetmore

United States
Environmental Protection
Agency

Heuristics for Chemical Use

Chemical Use Categories estimated from ACToR (computational toxicology database):

- The sources for chemical data were assigned to various chemical use categories.
- Chemicals from multiple sources were assigned to multiple categories.

Table: Hits per use category for a given chemical

CASRN	Category 1	Category 2	…	Category 12
65277-42-1	0	10	…	1
50-41-9	31	7	…	3
…	…	…	…	…

Binary matrix

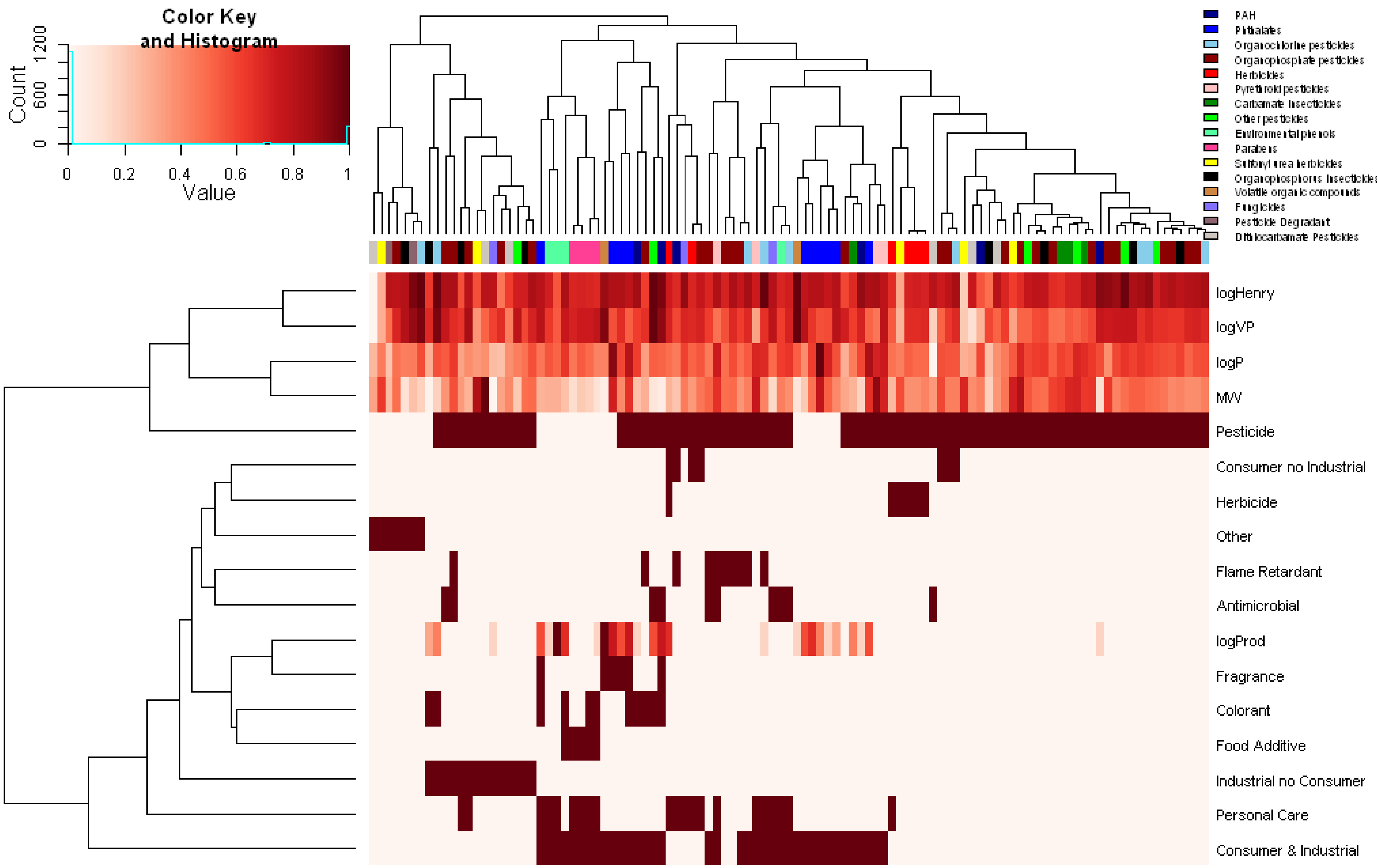
CASRN	Category 1	Category 2	…	Category 12
65277-42-1	0	1	…	0
50-41-9	1	1	…	0
…	…	…	…	…

12 Chemical Use Categories
Antimicrobials
Chemical Industrial Process
Consumer
Dyes and Colorants
Fertilizers
Food Additive
Fragrances
Herbicides
Personal Care Products
Pesticides
Petrochemicals
Other

Work by Alicia Frame, Kathie Dionisio, Richard Judson
Dionisio et al. *submitted*

United States Environmental Protection Agency

Heuristics for Chemical Use

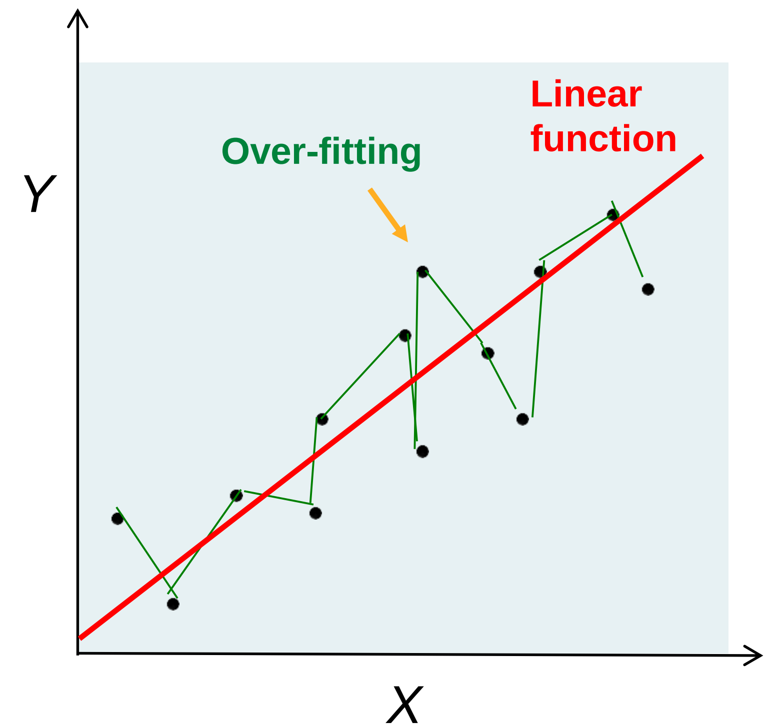


NHANES Chemicals

Wambaugh et al., submitted

Using Akaike Information Criterion (AIC) to Avoid Over-fitting

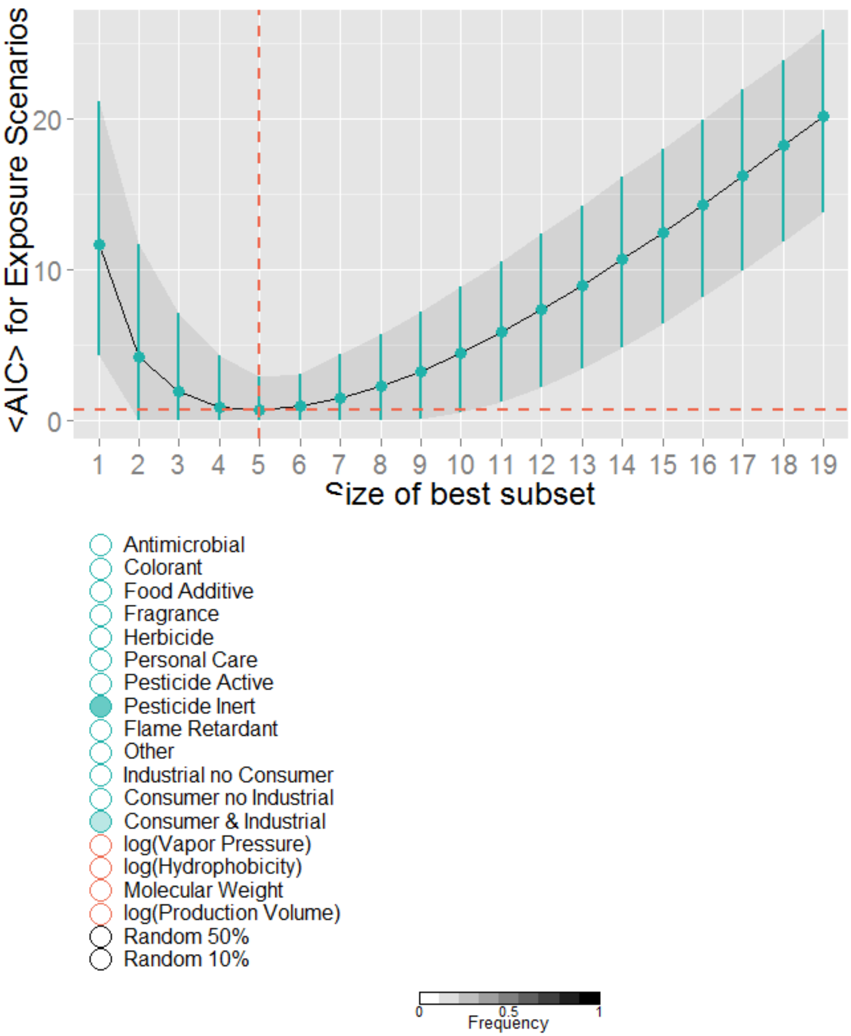
- In first generation analysis we were evaluating existing models with the available (few) chemicals
- Now we are trying to build a model using essentially the same number of chemicals: ***there is a danger of over-fitting***
- AIC (Akaike, 1974) is a statistical measure of model
    - Penalize goodness of fit with number of parameters needed to achieve that fit
    - Builds confidence that a model is appropriate for extrapolation to other circumstances
    - Model does not just describe the noise in a particular data set



Noisy data and of Over-fitting

United States
Environmental Protection
Agency

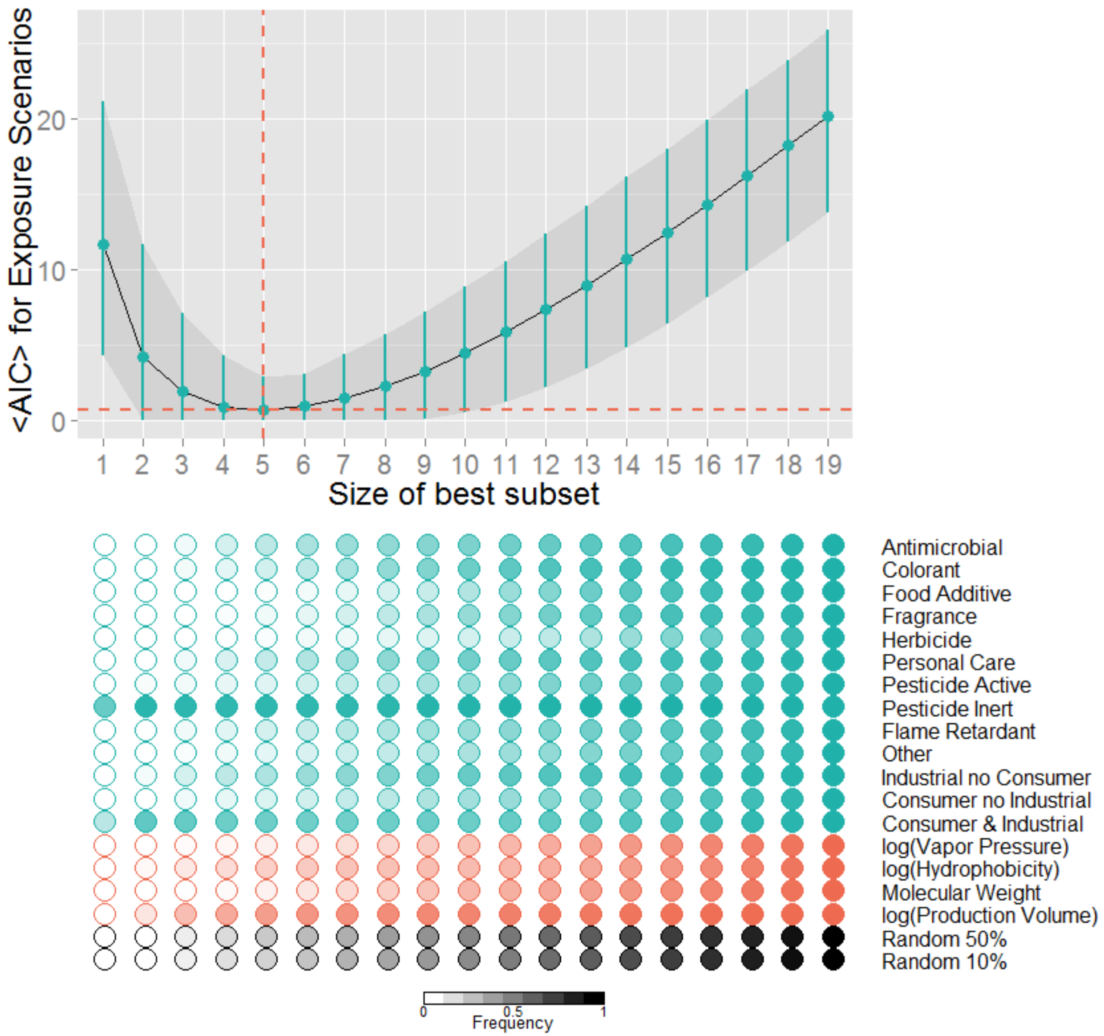
Not All Descriptors Are Useful



- The average relative AIC (smaller is better) for models made with different numbers of parameters for explaining 1500 different combinations of chemical exposures

- The predictors involved in the optimal model with higher frequencies are represented by darker circles, and those with lower frequencies by lighter circles

Not All Descriptors Are Useful



- The average relative AIC (smaller is better) for models made with different numbers of parameters for explaining 1500 different combinations of chemical exposures
- The predictors involved in the optimal model with higher frequencies are represented by darker circles, and those with lower frequencies by lighter circles
- As a sanity check, two random variables generated from binomial distribution with probability 50% and 10% of obtaining 1, are not selected as optimal descriptors in the five factor model

High-throughput exposure heuristics

Heuristic	Description	Number of Chemicals: Inferred NHANES Chemical Exposures (106)	Number of Chemicals: Full Chemical Library (7784)
ACToR "Consumer use & Chemical/Industrial Process use"	Chemical substances in consumer products (*e.g.*, toys, personal care products, clothes, furniture, and home-care products) that are also used in industrial manufacturing processes. Does not include food or pharmaceuticals.	37	683
ACToR "Chemical/Industrial Process use with no Consumer use"	Chemical substances and products in industrial manufacturing processes that are not used in consumer products. Does not include food or pharmaceuticals	14	282
ACToR UseDB "Pesticide Inert use"	Secondary (*i.e.*, non-active) ingredients in a pesticide which serve a purpose other than repelling pests. Pesticide use of these ingredients is known due to more stringent reporting standards for pesticide ingredients, but many of these chemicals appear to be also used in consumer products	16	816
ACToR "Pesticide Active use"	Active ingredients in products designed to prevent, destroy, repel, or reduce pests (*e.g.*, insect repellants, weed killers, and disinfectants).	76	877
TSCA IUR 2006 Total Production Volume	Sum total (kg/year) of production of the chemical from all sites that produced the chemical in quantities of 25,000 pounds or more per year. If information for a chemical is not available, it is assumed to be produced at <25,000 pounds per year.	106	7784

United States
Environmental Protection
Agency

NHANES Data Breaks Down by Demographics

Urinary Bisphenol A (2,2-*bis*[4-Hydroxyphenyl] propane)

Geometric mean and selected percentiles of urine concentrations (in µg/L) for the U.S.
and Nutrition Examination Survey.

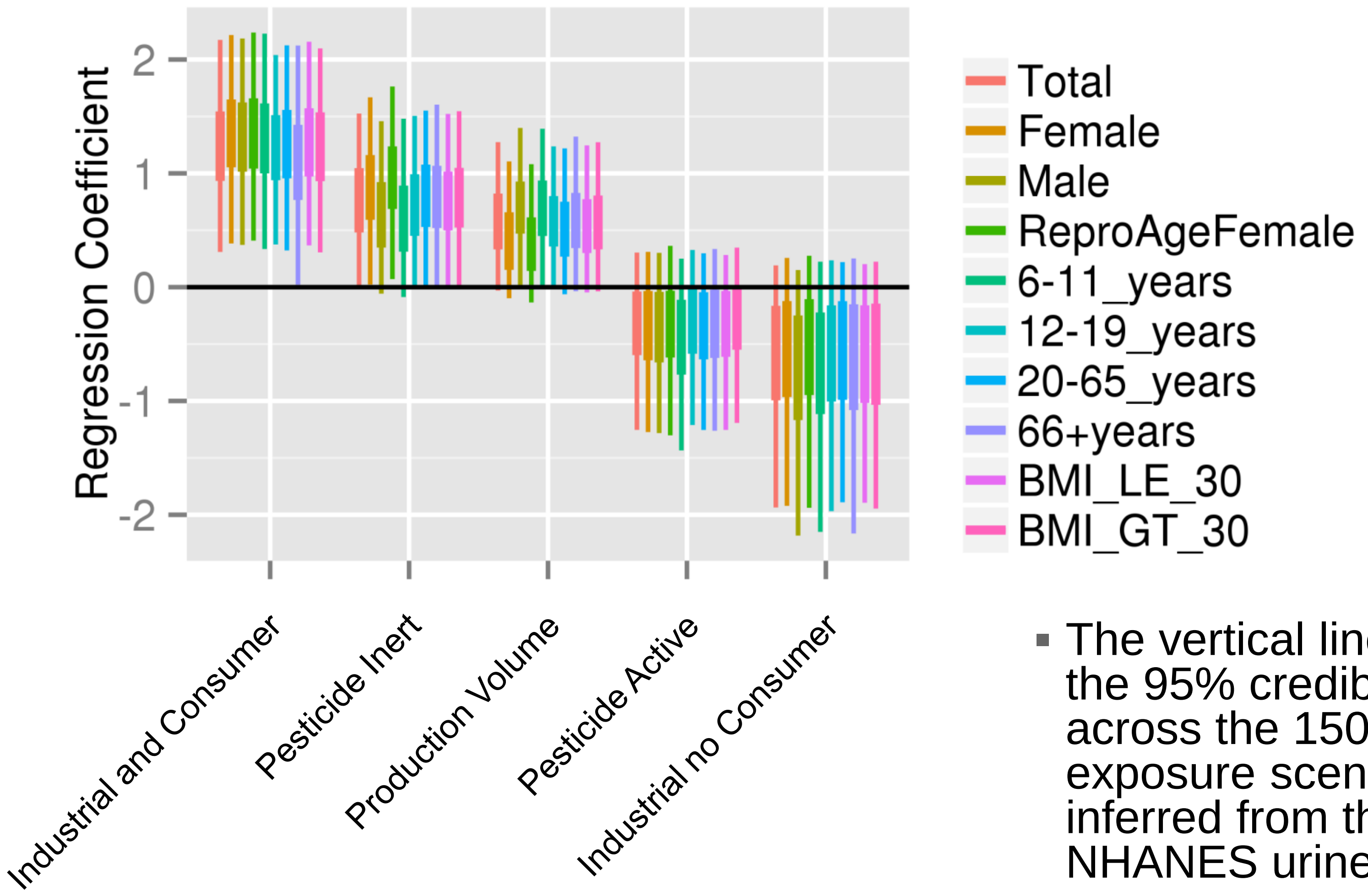
	Survey years	Geometric mean (95% conf. interval)	Selected pe (95% confiden 50th	75th
Total	03-04	**2.64** (2.38-2.94)	**2.80** (2.50-3.10)	**5.50** (5.00-6.20)
	05-06	**1.90** (1.79-2.02)	**2.00** (1.90-2.00)	**3.70** (3.50-3.90)
	07-08	**2.08** (1.92-2.26)	**2.10** (1.90-2.30)	**4.10** (3.60-4.60)
Age group				
6-11 years	03-04	**3.55** (2.95-4.29)	**3.80** (2.70-5.00)	**6.90** (6.00-8.30)
	05-06	**2.86** (2.52-3.24)	**2.70** (2.30-2.90)	**5.00** (4.40-5.80)
	07-08	**2.46** (2.20-2.75)	**2.40** (1.90-3.00)	**4.50** (3.70-5.50)
12-19 years	03-04	**3.74** (3.31-4.22)	**4.30** (3.60-4.60)	**7.80** (6.50-9.00)
	05-06	**2.42** (2.18-2.68)	**2.40** (2.10-2.70)	**4.30** (3.90-5.20)
	07-08	**2.44** (2.14-2.78)	**2.30** (2.10-2.60)	**4.40** (3.70-5.50)
20 years and older	03-04	**2.41** (2.15-2.72)	**2.60** (2.30-2.80)	**5.10** (4.50-5.70)
	05-06	**1.75** (1.62-1.89)	**1.80** (1.70-2.00)	**3.40** (3.10-3.70)
	07-08	**1.99** (1.82-2.18)	**2.00** (1.80-2.30)	**3.90** (3.40-4.60)

CDC, Fourth National Exposure Report (2011)

- Will different demographics have different heuristics?

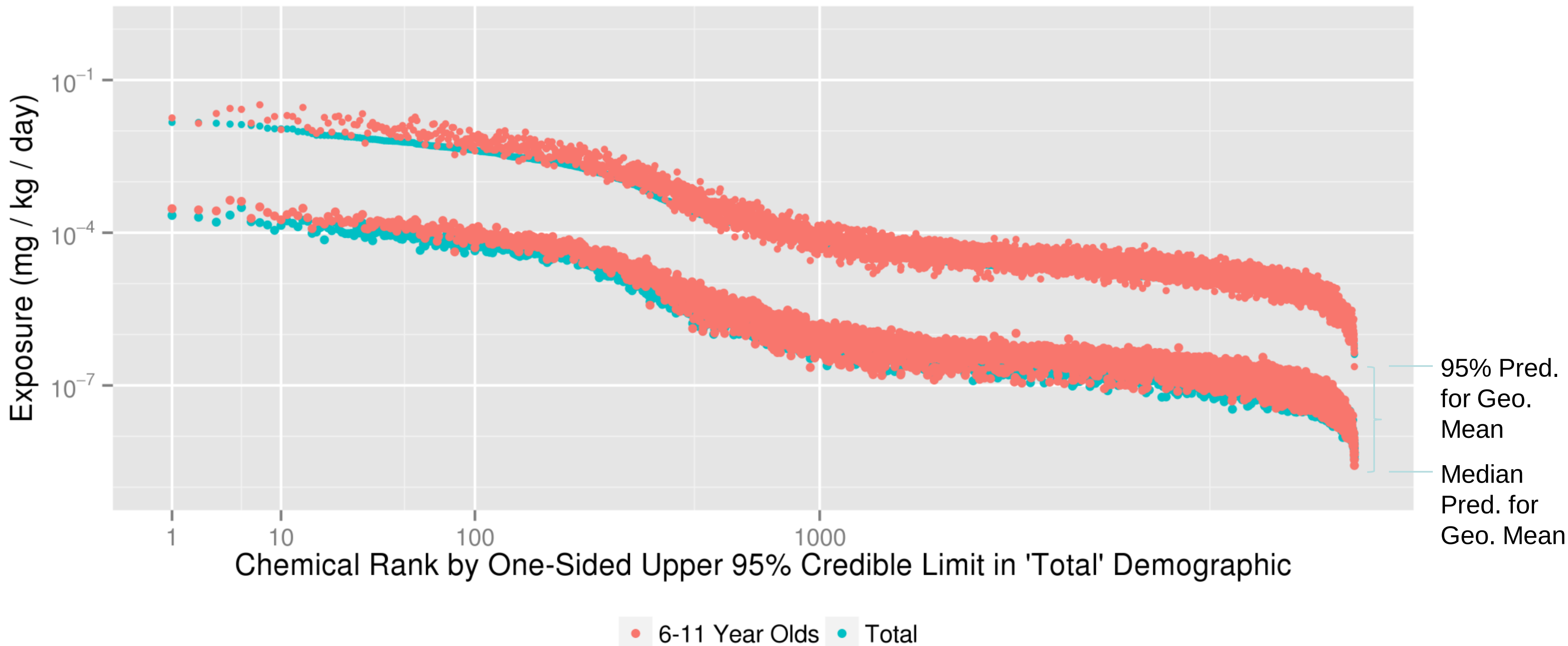
United States Environmental Protection Agency

Predictors Do Not Vary Between Groups



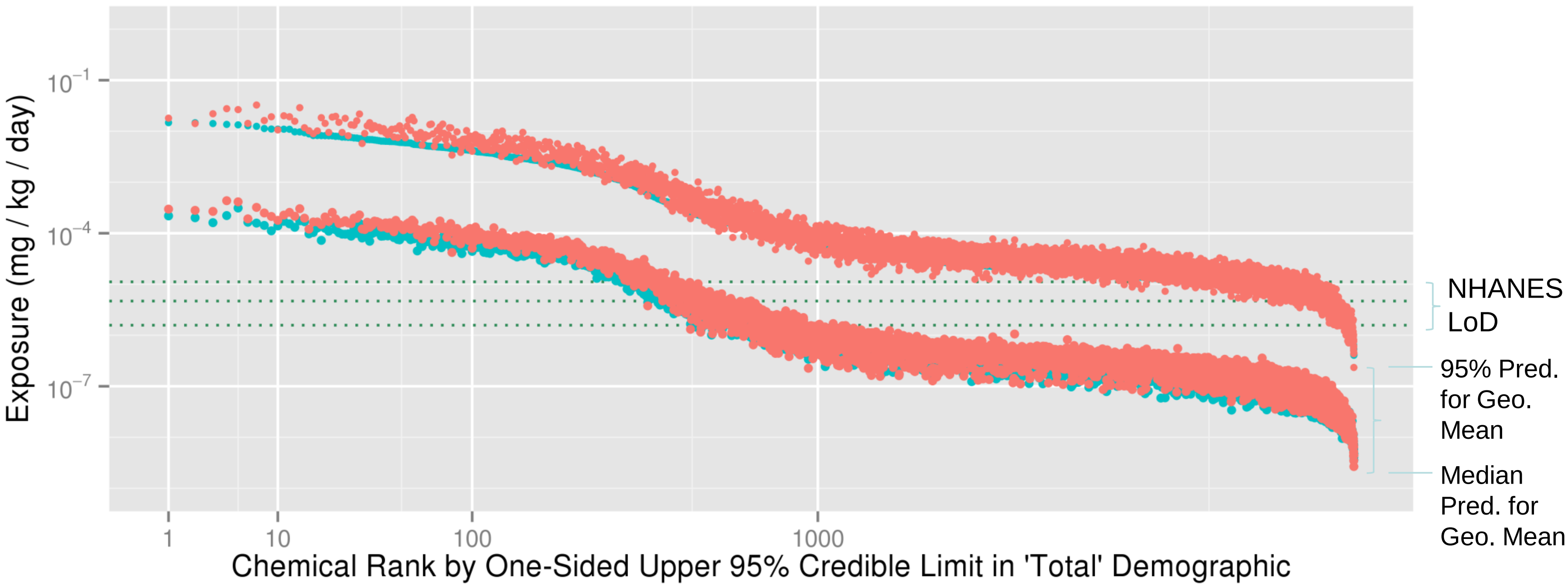
- The vertical lines indicate the 95% credible interval across the 1500 different exposure scenarios inferred from the NHANES urine data

Exposure Predictions for 7968 EDSP Chemicals



- These are the calibrated and evaluated predictions for the geometric mean exposure for the total NHANES population and for children aged 6-11

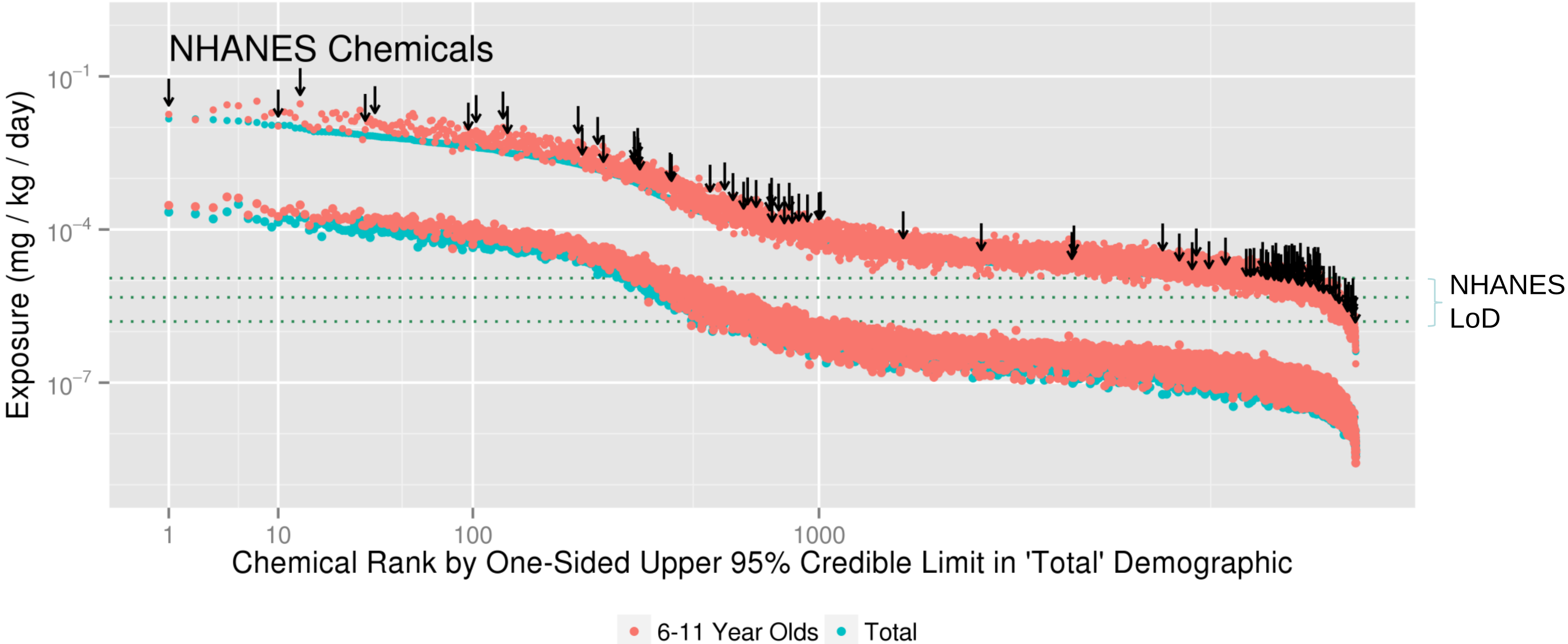
Exposure Predictions for 7968 EDSP Chemicals



- We focus on the median and upper 95% predictions because the lower 95% is below the NHANES limits of detection (LoD)
- Dotted lines indicate 25%, median, and 75% of the LoD distribution

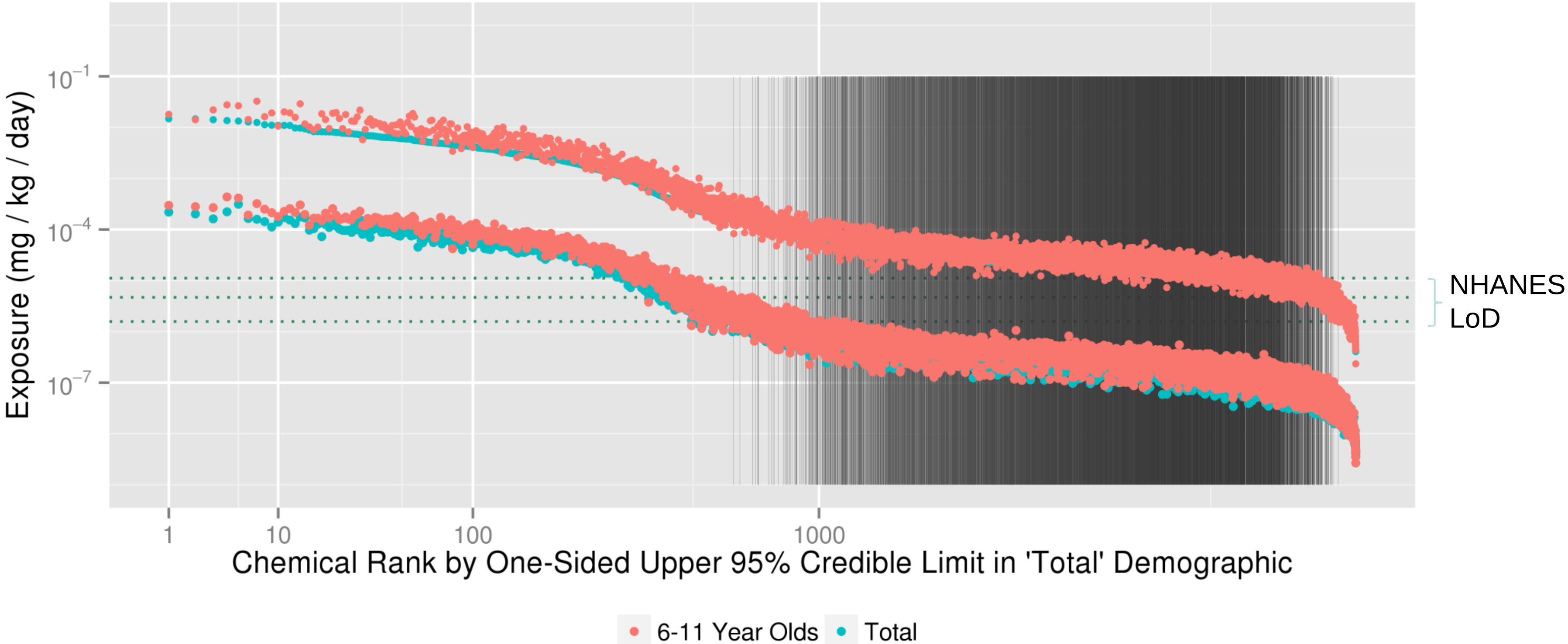
United States Environmental Protection Agency

Exposure Predictions for 7968 EDSP Chemicals



- Chemicals currently monitored by NHANES are distributed throughput the predictions
- Chemicals with the first and ninth highest 95% limit are monitored by NHANES

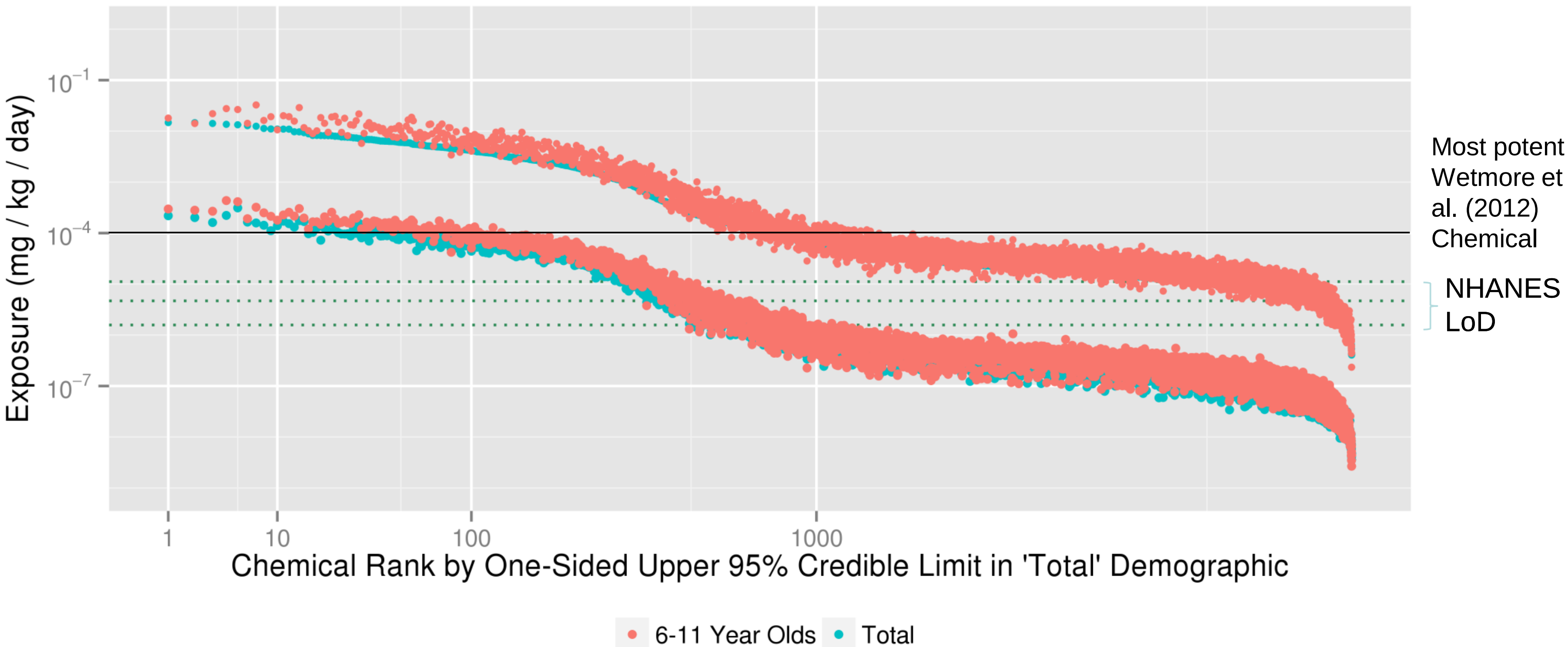
Exposure Predictions for 7968 EDSP Chemicals



- The grey stripes indicate the 4182 chemicals with no use indicated by ACToR UseDB for any of the four use category heuristics

United States Environmental Protection Agency

Exposure Predictions for 7968 EDSP Chemicals



- 95% confidence that the geometric mean exposures for many chemicalsare below the most potent bioactive dose for all 237 ToxCast chemicals examined in Wetmore et al. (2012)

Conclusions

- High throughput computational model predictions of exposure is possible
  - These prioritizations have been compared with CDC NHANES data, yielding empirical calibration and estimate of uncertainty
- Indoor/consumer use is a primary determinant of NHANES exposure (Wambaugh *et al*., 2013)
  - Developing and evaluating HT models for exposure from consumer use and indoor environment (*e.g.*, SHEDS-Lite from Isaacs *et al.*)
- Near field heuristics from new databases (*e.g.*, Goldsmith *et al.*, 2014) can explain a great deal of exposure from consumer products, but lack rapid exposure heuristics for articles and building materials

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."
Albert Einstein, quoted in J R Newman, *The World of Mathematics* (1956).

EPA Office of Research and Development

United States Environmental Protection Agency

ExpoCast Team

Kathie Dionisio*
Peter Eghehy
Kristin Isaacs
Richard Judson
Thomas Knudsen
Chantel Nicolas*
Robert Pearce*
James Rabinowitz
Woody Setzer
Cory Strope*
Dan Vallero
Amber Wang*
John Wambaugh

*Trainees

Chemical Safety for Sustainability (CSS) Collaborators

Michael Breen
Stephen Edwards
Rocky Goldsmith
Chris Grulke*
Haluk Ozkaynak
Jon Sobus
Mark Strynar
Cecilia Tan

Elaine Hubal
CSS Deputy National Program Director
Tina Bahadori
CSS National Program Director

External Collaborators
Jon Arnot (ARC)
Deborah Bennett (University of California, Irvine)
Alicia Frame (Dow Chemical)
Olivier Jolliet (University of Michigan)
Jade Mitchell (Michigan State)
Barbara Wetmore (Hamner)